

Evaluación automática y mejora de la prosodia en personas con síndrome de Down mediante el uso de juegos serios

Automatic evaluation and prosody improvement of people with Down syndrome based on serious games

Mario Corrales-Astorgano

Departamento de Informática, ECA-SIMM, Universidad de Valladolid
mcorrales@infor.uva.es

Resumen: El habla de las personas con síndrome de Down presenta una serie de problemas relacionados con la sintaxis, la semántica, la fonología o la pragmática. Más concretamente, la prosodia está también afectada, lo que puede producir problemas a esta población para su desarrollo personal y su integración social. Por ello, la práctica de esta competencia concreta es fundamental en la formación comunicativa de las personas con síndrome de Down. En este trabajo presentamos la utilización de un videojuego educativo orientado a la mejora de las habilidades comunicativas, especialmente la prosodia. Para ello, es necesario analizar las particularidades de esta población con el objetivo de diseñar el videojuego acorde a estas particularidades. Además, se incluye en este proyecto el análisis de las grabaciones obtenidas durante las sesiones de juego, con el objetivo de caracterizar las voces de las personas con síndrome de Down y con la idea de implementar una evaluación automática de dichas grabaciones.

Palabras clave: Síndrome de Down, prosodia, videojuegos educativos, juegos serios, caracterización del habla

Abstract: The speech of people with Down syndrome presents multiple disorders affecting the different components of language (syntax, semantics, phonology and pragmatics). In particular, prosody is also affected, conditioning their personal development and their social integration. Due to these difficulties, prosody training is fundamental in their speech therapy. In this work, the use of a video game focused on improving the communication skills of people with Down syndrome is presented. It is necessary to analyze the specific characteristics of this population with the aim of developing a video game adapted to this population. Moreover, an analysis of the recordings recorded during the game sessions is included in this work. The main aim of this analysis is the characterization of Down syndrome speech and an automatic evaluation of these recordings.

Keywords: Down syndrome, prosody, educational video games, serious games, speech characterization

1 Justificación de la investigación propuesta

Algunas personas con discapacidad intelectual (ID) tienen problemas en sus relaciones sociales debido a sus dificultades comunicativas (Chapman, 1997; Cleland et al., 2010; Martin et al., 2009). El habla en general (Kent y Vorperian, 2013) y la prosodia en particular (Stojanovik, 2011) están afectadas, lo que produce problemas en el control de algunas funciones comunicativas. Hay pocos trabajos en la literatura que hayan ana-

lizado el habla de personas con discapacidad intelectual utilizando un enfoque basado en la comparación entre diferentes corpus (Kent y Vorperian, 2013). La mayoría de los estudios en el estado del arte han seguido un enfoque basado en pruebas de percepción, no en la medición y comparación de variables acústicas extraídas directamente de los audio incluidos en un corpus. Existen dos razones principales que explican la escasez de estos trabajos: la falta de corpus específicos de personas con discapacidad intelectual y la di-

ficultad para analizar el habla de dichas personas. Con respecto a la falta de corpus específicos, la grabación de las voces de las personas con discapacidad intelectual en contextos controlados no es una tarea fácil debido a que estas personas presentan diferentes problemas cognitivos como problemas de memoria a corto plazo o déficit de atención, entre otros (Chapman, 1997).

En cuanto a las dificultades para analizar la alteración en el habla en personas con discapacidad intelectual, hay que tener en cuenta que este tipo de voz suele contener disfluencias y otros problemas de producción del habla originados por problemas fisiológicos. Como resultado, la calidad de la producción fonética es baja, lo que limita el uso de sistemas de reconocimiento automático de voz, por lo que la segmentación automática es, en muchos casos, inabordable (Feng et al., 2010). La prosodia opera a un nivel supra-segmental (por ejemplo, a nivel de palabras o frases), de manera que la extracción de las características prosódicas es un proceso más costoso. Por lo tanto, la posibilidad de analizar las producciones prosódicas de los usuarios en tiempo real y el entrenamiento de sistemas automáticos de identificación de problemas prosódicos proporcionan información útil para que los terapeutas que trabajan con personas con discapacidad intelectual puedan ayudar a estas personas a mejorar sus competencias lingüísticas.

Por otro lado, el uso de juegos serios para mejorar la motivación de los estudiantes en educación está ampliamente estudiado (McFarlane, Sparrowhawk, y Heald, 2002). Sin embargo, no existen apenas referencias sobre las ventajas de utilizarlas para mejorar la producción del habla y las habilidades prosódicas de las personas con síndrome de Down (Kent y Vorperian, 2013). Aunque existen algunas herramientas (Saz et al., 2009), no son ampliamente utilizadas por personas con síndrome de Down, ya que se necesita un alto grado de motivación, por lo que no es válido para este tipo de usuarios, debido a los problemas cognitivos que presentan.

2 Antecedentes y trabajo relacionado

Para analizar los problemas en la prosodia de las personas con síndrome de Down se han seguido principalmente dos enfoques: análisis

acústicos y análisis perceptuales. En ambos casos, es importante diferenciar entre adultos y niños, ya que las diferencias fisiológicas pueden influir en los resultados obtenidos. Con respecto a los análisis acústicos del habla, los estudios indican niveles más altos en la frecuencia fundamental (F0) en adultos con síndrome de Down (Lee, Thorpe, y Verhoeven, 2009; Albertini et al., 2010) en comparación con adultos sin ninguna discapacidad. Además, los adultos con síndrome de Down presentan menores perturbaciones en la frecuencia (*jitter*) (Lee, Thorpe, y Verhoeven, 2009; Seifpanahi, Bakhtiar, y Salmalian, 2011). En relación con la energía, se han observado valores más bajos en adultos con síndrome de Down y menores perturbaciones (*shimmer*) (Albertini et al., 2010). Las personas con síndrome de Down también presentan algunos problemas en la fluidez del habla, como el tartamudeo. Aunque los problemas de fluidez no son una característica universal en las personas con síndrome de Down, se ha demostrado que son un problema común en esta población (Van Borsel y Vandermeulen, 2008; Devenny y Silverman, 1990; Eggers y Van Eerdenbrugh, 2017).

Por otro lado, también existen resultados en la literatura en relación con niños con síndrome de Down en comparación con niños sin discapacidad intelectual. Se han observado valores más bajos en la frecuencia fundamental (Zampini et al., 2016), perturbaciones en la frecuencia más altas (Moura et al., 2008) y perturbaciones en la energía más altas (Moura et al., 2008).

Con respecto a los análisis perceptuales, algunos estudios catalogan la voz de los adultos con síndrome de Down como ronca (Moran y Gilbert, 1982). Además, también se han observado discrepancias entre los resultados de los juicios perceptuales en comparación con los análisis acústicos, concretamente en lo relacionado con la frecuencia (Rodger, 2009).

En relación con el uso de videojuegos en educación especial, existen numerosos estudios que demuestran la eficacia del uso de videojuegos para la rehabilitación y entrenamiento de personas con discapacidad intelectual. Algunos ejemplos son la mejora en el tiempo de reacción (Standen et al., 2009), el entrenamiento en la toma de decisiones (Standen, Rees, y Brown, 2009) o el entrenamiento de las habilidades matemáticas (Shafie et al., 2013).

3 Descripción de la investigación propuesta

Se propone un trabajo de investigación orientado al desarrollo de un videojuego para la mejora de las habilidades comunicativas de personas con síndrome de Down, especialmente enfocado en la prosodia. Para ello, lo primero fue realizar un estudio de los principales problemas que presenta esta población en relación con sus habilidades comunicativas y el desarrollo de actividades orientadas a mejorar estos aspectos. En paralelo, fue importante definir los escenarios donde estas actividades están incluidas, así como la historia en la que se desarrollan dichas actividades. Además, fue necesario realizar un análisis conciso sobre como los problemas de las personas con síndrome de Down podían afectar a su interacción con el videojuego.

El género elegido para implementar el videojuego es un aspecto clave. En este caso, hemos utilizado el género de las aventuras gráficas, ya que permite incluir dentro de una narrativa una serie de actividades diseñadas para mejorar las habilidades comunicativas. El videojuego incluye conversaciones con personajes, uso de objetos y navegación por diferentes escenarios. Los jugadores tienen que usar el ratón del ordenador para interactuar con los diferentes elementos del escenario. La principal innovación de nuestra propuesta con respecto a otras desarrolladas es que el jugador tiene que reproducir diferentes frases durante el transcurso del juego y estas reproducciones son almacenadas, por lo cual pueden ser analizadas posteriormente. Al lado del jugador, siempre es necesario un acompañante que evalúe en tiempo real si estas grabaciones son lo suficiente buenas como para continuar con el juego o, por el contrario, el jugador tiene que repetir la actividad. Esta persona también ejerce el rol de apoyo al jugador en el caso que se quede bloqueado o de motivador extra para que el jugador continúe con el videojuego.

El videojuego incluye actividades que introducen frases de diferentes modalidades (declarativas, interrogativas y exclamativas). Normalmente, los patrones de entonación varían dependiendo de la modalidad de la frase. Las frases declarativas normalmente acaban con una bajada hacia tonos bajos, mientras que las interrogativas acaban con una subida hacia tonos altos. Por otro lado, las frases parcialmente interrogativas, que están



Figura 1: Escenario del juego con una actividad de producción

caracterizadas por un elemento interrogativo al principio de la frase, empiezan con tonos altos asociados a la parte interrogativa y acaban con una bajada hacia tonos bajos. Las frases exclamativas son, normalmente, variaciones de las frases declarativas, por lo que la variación radica en aspectos como la energía, el volumen o la modulación del tono usado por el hablante. Además, la combinación de diferentes frases con diferentes modalidades permite la inclusión de inflexiones que indican una segmentación en la producción oral. Dependiendo del contexto y de la velocidad de producción de la frase, estas inflexiones pueden corresponder a una pausa (silencio) y a un fin de frase, o a una semi-pausa, que implica un cambio de entonación dentro de la misma frase.

4 Metodología y experimentos propuestos

En el estado actual del trabajo de investigación se han realizado dos experimentos cuyos resultados han sido publicados. Por un lado, para analizar la interacción entre el jugador y el videojuego, se realizaron una serie de tests de usabilidad combinados con los datos sobre la sesión de juego que el propio videojuego almacena automáticamente (González-Ferreras et al., 2017). Por otro lado, se realizó un estudio de las grabaciones obtenidas durante las sesiones de juego para evaluar las diferencias entre las grabaciones realizadas por personas con síndrome de Down y las realizadas por personas sin discapacidad intelectual.

Con respecto al análisis de las grabaciones obtenidas en las diferentes sesiones de juego, el objetivo es identificar las características acústicas que caracterizan el habla de las

personas con síndrome de Down. Estas características están relacionadas con los dominios de la frecuencia, de la energía, temporal y espectral. Para realizar este proceso, se llevaron a cabo unas grabaciones de las mismas frases del juego, pero grabadas por personas sin discapacidad intelectual. Para ello, se extrajeron una serie de características pertenecientes a cada dominio estudiado y se compararon estas características entre las extraídas de grabaciones realizadas por personas con síndrome de Down y las extraídas de grabaciones realizadas por personas sin discapacidad intelectual. De esta comparación surgieron las características con mayor capacidad de discriminación entre los dos grupos, utilizando para ello tests estadísticos. Además, se entrenaron tres clasificadores binarios utilizando las características significativamente diferentes y se realizó una comparación entre las características de los diferentes dominios para analizar cual era más relevante a la hora de clasificar una voz como proveniente de una persona con síndrome de Down. Finalmente, se realizó un test perceptual utilizando grabaciones creadas utilizando un algoritmo de transferencia de prosodia entre los grupos: la prosodia de las grabaciones de uno de los grupos se transfirió a las grabaciones del otro grupo, y viceversa. Los resultados de este test mostraron la importancia de la entonación y el ritmo a la hora de identificar un habla como atípica (Corrales-Astorgano, Escudero-Mancebo, y González-Ferreras, 2018).

El desarrollo de un módulo de evaluación automática de las grabaciones realizadas por los jugadores se desarrollará en dos etapas. En una primera etapa se monitorizará el trabajo de usuario y entrenador en las actividades de producción oral integradas en el videojuego. Para ello, se desarrollará una interfaz que permita recoger no sólo las respuestas orales de los usuarios y las puntuaciones asignadas por el entrenador, sino también las reacciones de este último en forma de ayuda, orientación, etc. Una segunda etapa de análisis de las muestras capturadas permitirá validar una serie de informes de calidad que identifiquen cuáles son las variables acústico/prosódicas más relevantes a la hora de determinar la calidad de los turnos de los usuarios. Con dicha información, se entrenará un sistema de predicción automática de calidad que sea capaz de valorar la respuesta oral del usuario a la vez que ofrecer información para seguir me-

jorando (feedback).

5 Cuestiones de investigación

Debido al momento concreto en el que se encuentra la investigación, a parte de discutir las metodologías y resultados obtenidos en los estudios presentados en la anterior sección, las cuestiones a discutir podrían ir enfocadas en como conseguir desarrollar el módulo de evaluación automática. Algunas cuestiones concretas podrían ser:

- ¿Es posible correlar las evaluaciones de las grabaciones realizadas por los profesores con las características acústicas extraídas de las mismas?
- ¿Son eficaces los clasificadores binarios para realizar esta tarea?. Y si es así, ¿cómo mejorar los resultados de clasificación?
- ¿Es posible realizar este proceso en tiempo real o sería mejor continuar utilizando al profesor como ayudante y utilizar el módulo como evaluador complementario?

Agradecimientos

Este trabajo es financiado parcialmente por el proyecto del Ministerio de Economía, Industria y Competitividad con título “INCORPORACIÓN DE UN MÓDULO DE PREDICCIÓN AUTOMÁTICA DE LA CALIDAD DE LA COMUNICACIÓN ORAL DE PERSONAS CON SÍNDROME DE DOWN EN UN VIDEOJUEGO EDUCATIVO”, clave TIN2017-88858-C2-1-R.

Bibliografía

- Albertini, G., S. Bonassi, V. Dall’Armi, I. Giachetti, S. Giaquinto, y M. Mignano. 2010. Spectral analysis of the voice in Down syndrome. *Research in developmental disabilities*, 31(5):995–1001.
- Chapman, R. S. 1997. Language development in children and adolescents with Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, 3(4):307–312.
- Cleland, J., S. Wood, W. Hardcastle, J. Wishart, y C. Timmins. 2010. Relationship between speech, oromotor, language and cognitive abilities in children with Down’s syndrome. *International journal*

- of language & communication disorders, 45(1):83–95.
- Corrales-Astorgano, M., D. Escudero-Mancebo, y C. González-Ferreras. 2018. Acoustic characterization and perceptual analysis of the relative importance of prosody in speech of people with Down syndrome. *Speech Communication*, 99:90–100.
- Devenny, D. y W. Silverman. 1990. Speech dysfluency and manual specialization in Down’s syndrome. *Journal of Intellectual Disability Research*, 34(3):253–260.
- Eggers, K. y S. Van Eerdenbrugh. 2017. Speech disfluencies in children with Down Syndrome. *Journal of Communication Disorders*.
- Feng, J., J. Lazar, L. Kumin, y A. Ozok. 2010. Computer Usage by Children with Down Syndrome: Challenges and Future Research. *ACM Transactions on Accessible Computing*, 2(3):13.
- González-Ferreras, C., D. Escudero-Mancebo, M. Corrales-Astorgano, L. Aguilar-Cuevas, y V. Flores-Lucas. 2017. Engaging adolescents with Down syndrome in an educational video game. *International Journal of Human-Computer Interaction*, páginas 1–20.
- Kent, R. D. y H. K. Vorperian. 2013. Speech impairment in Down syndrome: a review. *Journal of Speech, Language, and Hearing Research*, 56(1):178–210.
- Lee, M. T., J. Thorpe, y J. Verhoeven. 2009. Intonation and phonation in young adults with Down syndrome. *Journal of Voice*, 23(1):82–87.
- Martin, G. E., J. Klusek, B. Estigarribia, y J. E. Roberts. 2009. Language characteristics of individuals with Down syndrome. *Topics in Language Disorders*, 29(2):112.
- McFarlane, A., A. Sparrowhawk, y Y. Heald. 2002. *Report on the educational use of games*. TEEM (Teachers evaluating educational multimedia), Cambridge.
- Moran, M. J. y H. R. Gilbert. 1982. Selected acoustic characteristics and listener judgments of the voice of Down syndrome adults. *American journal of mental deficiency*.
- Moura, C. P., L. M. Cunha, H. Vilarinho, M. J. Cunha, D. Freitas, M. Palha, S. M. Pueschel, y M. Pais-Clemente. 2008. Voice parameters in children with Down syndrome. *Journal of Voice*, 22(1):34–42.
- Rodger, R. 2009. *Voice quality of children and young people with Down’s Syndrome and its impact on listener judgement*. Ph.D. tesis, Queen Margaret University.
- Saz, O., S. Yin, E. Lleida, R. Rose, C. Vaqueiro, y W. R. Rodríguez. 2009. Tools and Technologies for Computer-Aided Speech and Language Therapy. *Speech Communication*, 51(10):948–967.
- Seifpanahi, S., M. Bakhtiar, y T. Salmalian. 2011. Objective vocal parameters in Farsi-speaking adults with Down syndrome. *Folia Phoniatrica et Logopaedica*, 63(2):72–76.
- Shafie, A., W. F. Wan Ahmad, N. Mohd, J. J. Barnachea, M. F. Taha, y R. L. Yusuff. 2013. “SynMax”: A mathematics application tool for down syndrome children. En *Advances in Visual Informatics*. Springer, páginas 615–626.
- Standen, P., N. Anderton, R. Karsandas, S. Battersby, y D. Brown. 2009. An evaluation of the use of a computer game in improving the choice reaction time of adults with intellectual disabilities. *Journal of Assistive Technologies*, 3(4):4–11.
- Standen, P., F. Rees, y D. Brown. 2009. Effect of playing computer games on decision making in people with intellectual disabilities. *Journal of Assistive Technologies*, 3(2):4–12.
- Stojanovik, V. 2011. Prosodic deficits in children with Down syndrome. *Journal of Neurolinguistics*, 24(2):145–155.
- Van Borsel, J. y A. Vandermeulen. 2008. Cluttering in Down syndrome. *Folia Phoniatrica et Logopaedica*, 60(6):312–317.
- Zampini, L., M. Fasolo, M. Spinelli, P. Zanchi, C. Suttora, y N. Salerni. 2016. Prosodic skills in children with Down syndrome and in typically developing children. *International Journal of Language & Communication Disorders*, 51(1):74–83.